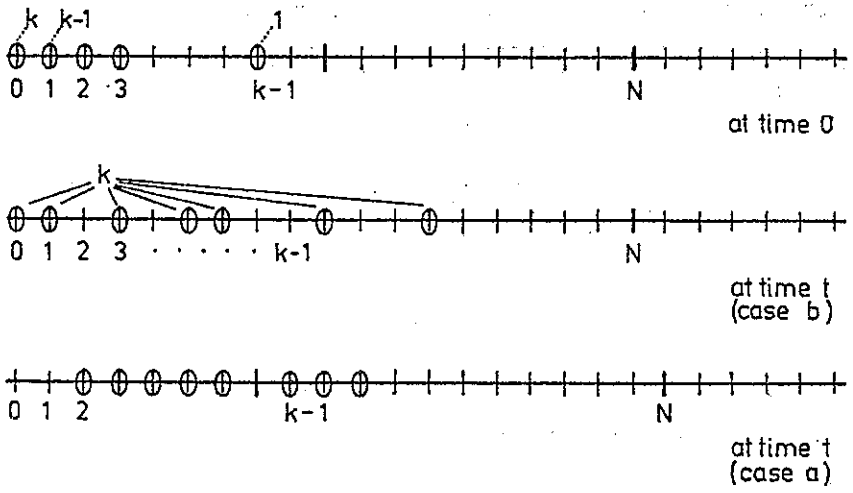
(75)

# ANALYTICAL METHODS IN DISTRIBUTED SYSTEMS*

*M. Arató*

In a network transmission line there are two possibilities to send a message of k packets from node 0 to node N /on a real line/. In the first case a) all the packets are moving together this is the message switching case, and in the other case b) the packets are moving independently /if one can




do it/, this is the packet switching case. After t moments one may have the situation which is described on Fig.1. For simplicity let us say that we have k balls /at time moment t=o/ at the points 0, 1, ..., k-1 and this set of balls represents the message. Each ball has the possibility to

*The paper is the version of the author's lecture on the 5-th Visegrád Winter School on Operating Systems (1979, 28-30 January).

move right with probability 1-p and to remain on the same place with probability p. In the first model /model a)/ the balls are connected, which means they can move only together, and the probability that the message makes 1 step is

$$q_k = (1-p)^k$$

and the probability that after time t the message makes i steps is

$$\binom{t}{i} q_k^i (1-q_k)^{t-i}.$$

In the second model /model b)/ a ball on place j will stay with probability p /under condition that it is possible to move/, or will stay with probability 1 /under condition it is impossible to move, because the place j+1 is occupied/.

Let $\tau$ denote the time when the k th ball /which is at time moment t=o in O/ will be in N. What is the expectation $E\tau$ in case a) and b)? It is easy to prove that in case a) we have

$$E_a \tau = \sum_{i=0}^{\infty} (N+i) \binom{N+i}{i} q_k^N (1-q_k)^i = N+(N+1) \frac{1-q_k}{q_k} =$$

$$= \frac{N+1 - q_k}{q_k}.$$

On the other side it can be easily proved that

$$E_b \tau \leq E_a \tau$$

but the exact value of $E_b \tau$ is not known. This exercise is only a simple example for analytical methods in distributed systems. In the later we shall discuss some other problems of distributed systems.

In the preceding we considered a network transmission path with N+1 nodes and N channels which were assumed to be noiseless, perfectly reliable and to have unit capacity. We assume that the transmission path is a part

of a larger network and so some other traffic can interfere with transmission. This fact is taken into consideration that each node has the same probability 1-p to be free, independently of each other and of their past. This is called a tandem net with $N$ nodes.

In the message switching case it may be assumed that once a node becomes free the message reserves it for all the transmission time. As the channel capacities are unit the transmission time for message of length $k$ at a node is a random variable $\xi$ with expectation $E\xi = k-1+(1-p)^{-1} = \mu_k$ and variance $D^2\xi = \frac{p}{(1-p)}$. If the message should wait for $k$ consecutive "free" states of the node then

$$(1) \qquad E\xi = \sum_{i=1}^{k} (1-p)^i, \quad D^2\xi = \frac{p}{(1-p)^k} \sum_{i=1}^{k} (1-p)^i \mu_i.$$

If numbered packets are sent successively through the network they move with different speed rates, and the packet having the greater number needs the longer time for travelling through the path. The packets with greater number have often to wait as packets with smaller number are occupying the nodes.

We are interested in the message delay in both cases, i.e. in the message switching and packet switching cases. Let $D_{k,n}$ denote the message delay in the message switching case, then

$$D_{kn} = \xi_1 + \xi_2 + \ldots + \xi_n$$

where the random variables $\xi_i$ are independent, identically distributed. Let $\bar{D}_{k,n}$ denote the message delay time in packet switching case, and $\xi_{ij}$ the time spent by the $i$ th packet at the $j$ th node after $(i-1)$ th packet has passed it. The $\xi_{ij}$ variables are independent with geometrical distribution and

$$\bar{D}_{1,j} = \xi_{11} + \xi_{12} + \ldots + \xi_{1j},$$

$$\bar{D}_{i,j} = \max (\bar{D}_{i,j-1}, \bar{D}_{i-1,j}) + \xi_{ij}, \quad (D_{i,o} = 0).$$

It is conjectured, that

$$\lim_{n\to\infty} \frac{\bar{D}_{k,n}}{n} = \frac{1}{1-p} ,$$

and the variance, distribution can also be calculated.*

From the above definition we easily get that

$$\sum_{j=1}^{n} \xi_{1,j} + \sum_{i=2}^{k} \xi_{i,n} \le \bar{D}_{k,n}$$

and as

$$\frac{1}{n} \sum_{j=1}^{n} \xi_{1,j} \to \frac{1}{1-p}$$

$$\frac{1}{1-p} \le \frac{1}{n} \sum_{j=1}^{n} \xi_{1,j} + \frac{1}{n} \sum_{i=2}^{k} \xi_{i,n} \le \liminf_{n\to\infty} \frac{1}{n} \bar{D}_{k,n}.$$

## 1. ON DISTRIBUTED SYSTEMS

A growing interest exists in distributed systems, and there is little doubt that distributed computing will be the primary systems architecture of the future. This means that systems are evolving toward a greater diffusion of processing power and databases.

The advantages of decentralised computing can be further enhanced by linking the distributed parts through electronic communications. We distinguish

a) centralized systems,
b) decentralized systems,
c) variations on the centralized/decentralized structures.

The main features of distributed systems are the following:

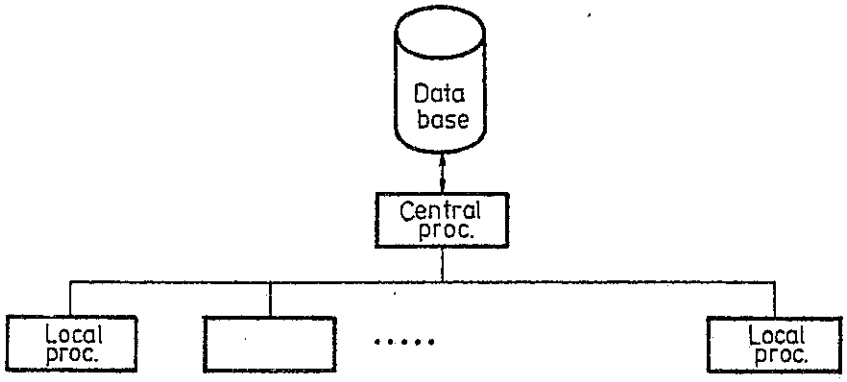
- multiple processors having general purpose computing compatibilities,
- communication links /often only intermittent/ among processors,

* This problem vas partially solved in a recent paper by T.Móri.

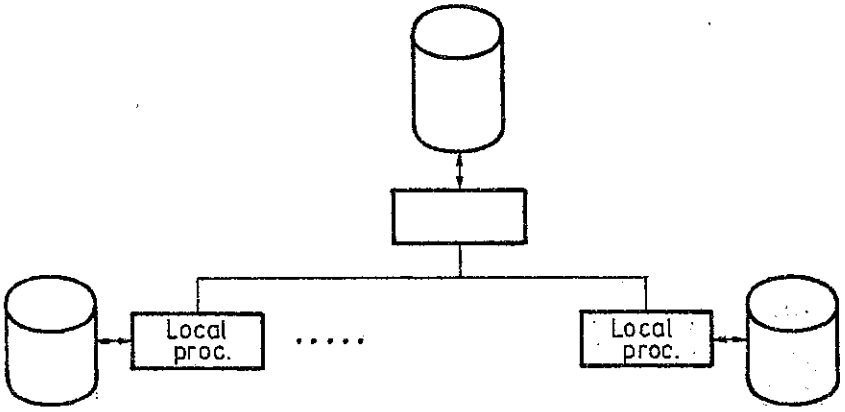
- relatively weak interactions among the distributed subsystems,
- considerable centralized coordination in the design and operation of the separate processing subsystems.

Three different version of distributed systems we give as alternative configurations as follows:

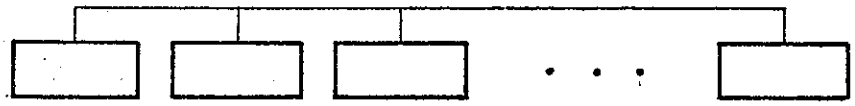
1. Distributed processing without local database




2. Hierarchical systems with non-shared local databases



3. Fully distributed network



There are many arguments on centralization or decentralization some of them we recount here:

1. *Centralization* /advantages and disadvantages/

- hardware economies of scale,
- operating economies of scale,
- more powerful capabilities,
- demand smoothing,
- reduced number of facilities,
- development of professional staff,
- increased system integration.

2. *Decentralization*

- greater control by users,
- increased motivation and involvment of users,
- economies of specialization,
- permits exploitation of low-cost minicomputers,
- permits small increments to capacity,
- reduces communication costs,
- reduced interactive response time,
- increased reliability,
- increased predictability of costs.

Distributed systems may be viewed as a hybrid between centralisation and decentralisation. Here we recall only:

| | |
|---|---|
| Advantages | economies of scale,<br>increased efficiency,<br>demand smoothing,<br>incorporation /user - technical expertise/,<br>integration of information processing,<br>integration of organizational activities,<br>greater user control,<br>reduced communication cost,<br>reduced response time,<br>reliability. |
| Hazards | creeping escalation,<br>hidden costs,<br>duplication and incompatibility,<br>incompetent design,<br>suboptimization. |

Distributed systems are not simple extensions of monolithic systems.

There are computer systems /special purpose or general/ which are operating in a noncentralized manner. Packet-switching computer communication networks are examples of distributed systems. In such systems two kinds of resources must be considered:

- system resources /multiplexing is required/,
- user resources.

An example for the second is distributed data-base sharing.

Among the many factors that motivate the interest to distributed systems are the desire to share resources and the need to achieve higher system performance and reliability.

Computer systems carry out two different kinds of activities:

- operation activities /compilers, assamblers, programs/,
- decision - making activities /Schedules, resource manager/.

M. Arató

Distributed systems are formed through a more or less loosecoupling of autonomous subsystems, each of which is assigned to perform a specific subtask of the global system task. Cooperation among such units is needed and each unit behaves as a simple computer installation.

In this paper we try to give some aspects of:

a) resource allocation problem, resource management, optimal file allocation,

b) scheduling, performance problems.

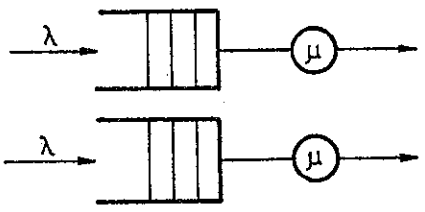
## 2. RESOURCE ALLOCATION PROBLEMS

One of the problems when we design distributed computer systems is the scheduling of jobs among processors in the system. The objective is to achive systeme balance, with a resultant performance increase, by automatically shifting jobs from heavily loaded processors to lightly loaded processors in the system.

Load balancing can be done statistically, or dynamically as the load and the state of the system changes. Dynamic load balancing have taken two different approaches. One is a combinatorial optimization problem. Jobes are reassigned dynamically by monitoring the state of the system. The other approach is to develop queuing models to analyzę the performance of the systems incorporating simple job routing policies that automatically shift jobs. In the following

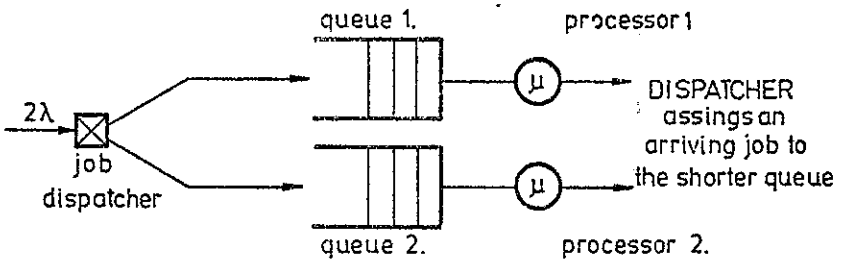
TWO PROCESSOR DISTRIBUTED COMPUTER SYSTEM
with load balancing

(1)



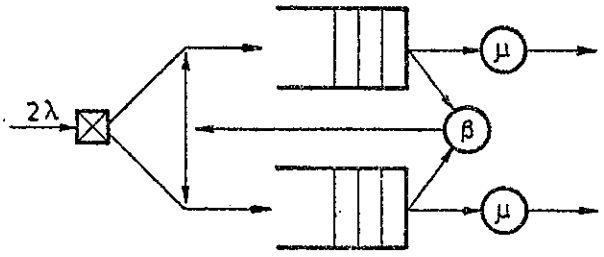

Join random queue without channel transfer

(2)



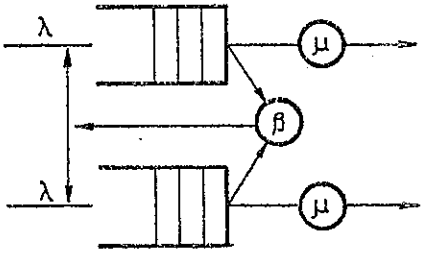

Join shorter queue without channel transfer

(3)



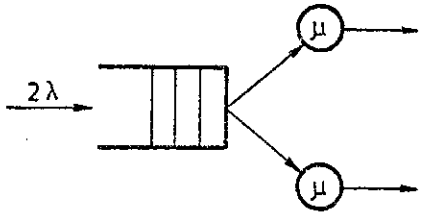

Join shorter queue with channel transfer

(4)



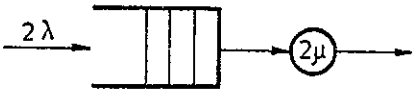

Join random queue with channel transfer

(5)




Instantaneous channel transfer

(6)




Single fast processor system

It is well known that the description of the system is given by the probability distributions

$$P_{ij}(t) = P\{X_1(t) = i,\ X_2(t) = j\}$$

/where $X_i(t)$ is the number of jobs in queue i/. The Kolmogorov equations can be written in the ordinary way for the $P_{ij}(t)$ and the steady state equations can be used for the following characteristics. Utilization of the processors: $U = \lambda/\mu$, average queue length (L), mean job turn around time (T) which by Little law equals to $\lambda T = L$.

In case $\mu = 1$ the following values of T can be given:

| Model \ λ | 0.1 | 0.2 | 0.5 | 0.9 |
|---|---|---|---|---|
| 1. | 1.11 | 1.25 | 2.00 | 10.00 |
| 2. | 1.02 | 1.06 | 1.42 | 5.47 |
| 3. | 1.02 | 1.05 | 1.39 | 5.35 |
| 5. | 1.01 | 1.04 | 1.33 | 5.26 |
| 6. | 0.56 | 0.63 | 1.00 | 5.00 |

## 3. RESOURCE MANAGEMENT PROBLEMS

In a single computer installation the resource management task is normally performed by the operating system and by the operator.

Relationship will exist between workload /input/ and computer performance, which depends on resource strategies. Resource management strategy allows intervention on some parameters, by which one can tune the system. Let us introduce the following symbols and definitions

- $w$ - computer workload,
- $y$ - performance, where $y = \underline{P}(m,w)$
- $a$ - parameters, $(a \in M)$,
- $g(\underline{a},w)$ - cost function.

Then in most cases we are interested in finding

$$g(\hat{a},w) = \inf_{a \in M} g(a,w).$$

The same resource management in computer network means further that

$y_i = P_i(a_i, w_i, u_i)$, where $u_i$ means interaction variable /e.g. amount of exchange workload at i/,

and

$g_i = g_i(a_i, w_i, u_i)$ would reflect network configuration and routing algorithmus $(i = 1,2,\ldots,n)$.

In this case it should be find

$$g(\hat{\underline{a}}, \underline{w}, \underline{u}) = \inf_{a \in M} (g\ \underline{a}, \underline{w}, \underline{u}).$$

This very formal description shows that there are two ways to overcome the conflict between global and local optimum control on resources. Entrusting the global resource management task to just one decision making unit within the network /„centralized approach"/. New decision-making unit /„coordinator"/ whose goal is hierarchically coordinate the local computer systems.

In most cases the computer workload, performance and parameters should be considered random variables or random vectors. Then the problem of minimization is sophisticated.

## 4. FILE ALLOCATION IN AN INFORMATION NETWORK

Transactions with the multiple-located file give rise the querry traffic /to a simple copy/, and $q_i$ means the volume of query, update traffic /to every copy/, and $r_i$ is the volume of update at the i-th node. Let

$$a_i = \begin{cases} 1 & \text{file is in i-th node,} \qquad i = 1,2,\ldots,n, \\ 0 & \text{otherwise.} \end{cases}$$

The costs consist of communication costs for querry and update and further of file storage costs. Let

$d_{jk}$ - communication cost between nodes j,k,

$\sigma_k$ - storage cost at node k,

I - the index set having a file copy.

The total cost in one moment is given by

$$C_{tot} = \sum_{j=1}^{n} \left[ \sum_{k \in I} r_j d_{jk} + q_j \min_{k \in I} d_{jk} \right] + \sum_{k \in I} \sigma_k \ .$$

Assuming that the references at each node form a reference string $\xi_{jt}$ $(j = 1,2,\ldots,n;\ t = 0,1,2,\ldots)$ with stochastic behavior the cost should be given by expectation.

## 5. HIERARCHICAL ROUTING AND FLOW CONTROL POLICY
(for packet switched networks)

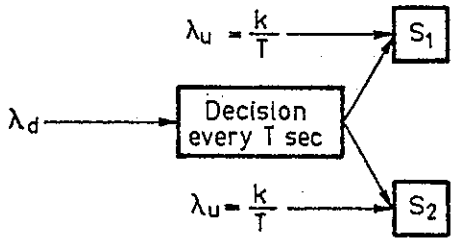
Management of traffic flow in a computer network can be divided into two main tasks, message routing and flow control.

Message delay or cost is a commonly used parameter to desingn routing policy and buffer size is the parameter for flow control algorithms.

It is desirable to study them jointly. Channel traffic intensity as a common parameter /it is closely related to delay and queue length/ is used to jointly optimize the routing and flow control.

Routing decisions in packet switched networks are often based upon congestion updates passed around the network. There is a natural conflict in the selection of the update period (T): make T small to have decisions on recent information, make T large to minimize the overhead.

Let us consider the following figure:




Poisson stream of data traffic with rate $\lambda_d$ is assumed. Decisions are made every T seconds as to which server to route to, based on which server had the smallest queue at the beginning of the T-period. All packets are routed to the chosen server for the entire T-period.

The overhead /updating the queue information every T seconds/ shall be simply modeled as an additional Poisson stream with rate $\lambda_u = \frac{k}{T}$.

All service times are mutually independent exponential with $\frac{1}{\mu}$. It is assumed

$$\lambda_d + 2\lambda_u < 2 \ .$$

The $\lambda_u$-streams represent update packets to accomplish node-to-node transfer of congestion information across the two links $S_1$ and $S_2$.

Other representations of the input process and the update overhead can be visualized /updates arrive deterministically, preemptive priority, different service times/.

REFERENCES

[1] M.ARATÓ: Statistical sequential methods in performance evaluation of computer systems. Modelling and performance evaluation of computer systems (1978) (Ed.-s: Beilner, Gelenbe) North Holland, 1-10.

[2] G.BUCCI - S.GOLINELLI: A distributed strategy for resource allocation in information networks. Int. Computing Symposium, 1977. 345-356.

[3] J.C.EMERY: Managerial and economic issues in distributed computing. Information Processing, 77. 945-955.

[4] C.HEWITT - M.BAKER: Actors and continous functionals. Formal Description of Programming Concepts, 1978. 357-390. N.M.P.C. 1978. Ed. E.J. Nenhold.

[5] L.KLEINROCK: Queuing systems. Vol.II. Computer Applications (1976), Wiley, New York.

[6] G.Le LANN: Distributed systems - toward a formal approach. Information Processing, 77. 455-460.

[7] M.MACKAWA: Interprocess communication in a highly diversified distributed systems. Information Processing, 77. 149-154.

[8] I.RUBIN: Message path delay in packet-switching communication networks. IEEE Trans on Communications 23. (1975), 186-192.

[9] F.E.TAYLOR: The relative merits of distributed computing systems. International Computing Symposium, 1977. 357-365. N.M.P., 1977. Ed. E. Morlet, D.Ribbens.

ÖSSZEFOGLALÁS

Osztott rendszerek analitikus problémái

*Arató Mátyás*

A dolgozatban a szerző egy hálózati üzenet továbbitási feladat kapcsán illusztrálja az analitikus kezdés nehézségeit. Az átlagos késési idő meghatározása különböző modellek esetén határeloszlástételek vizsgálatára vezethető vissza. Erőforrás elosztási és irányitási feladatok sorbanállási modellekkel oldhatók meg. A file elhelyezés és üzenet továbbitás problémaköre determinisztikus és sztochasztikus modellekkel is leirható.